Preface

An understanding of probability and statistics is an essential tool for a modern computer scientist. If your tastes run to theory, then you need to know a lot of probability (e.g., to understand randomized algorithms, to understand the probabilistic method in graph theory, to understand a lot of work on approximation, and so on) and at least enough statistics to bluff successfully on occasion. If your tastes run to the practical, you will find yourself constantly raiding the larder of statistical techniques (particularly classification, clustering, and regression). For example, much of modern artificial intelligence is built on clever pirating of statistical ideas. As another example, thinking about statistical inference for gigantic datasets has had a tremendous influence on how people build modern computer systems.

Computer science undergraduates traditionally are required to take either a course in probability, typically taught by the math department, or a course in statistics, typically taught by the statistics department. A curriculum committee in my department decided that the curricula of these courses could do with some revision. So I taught a trial version of a course, for which I wrote notes; these notes became this book. There is no new fact about probability or statistics here, but the selection of topics is my own; I think it's quite different from what one sees in other books.

The key principle in choosing what to write about was to cover the ideas in probability and statistics that I thought every computer science undergraduate student should have seen, whatever their chosen specialty or career. This means the book is broad and coverage of many areas is shallow. I think that's fine, because my purpose is to ensure that all have seen enough to know that, say, firing up a classification package will make many problems go away. So I've covered enough to get you started and to get you to realize that it's worth knowing more.

The notes I wrote have been useful to graduate students as well. In my experience, many learned some or all of this material without realizing how useful it was and then forgot it. If this happened to you, I hope the book is a stimulus to your memory. You really should have a grasp of all of this material. You might need to know more, but you certainly shouldn't know less.

Reading and Teaching This Book

I wrote this book to be taught, or read, by starting at the beginning and proceeding to the end. Different instructors or readers may have different needs, and so I sketch some pointers to what can be omitted below.

Describing Datasets

This part covers:

- Various descriptive statistics (mean, standard deviation, variance) and visualization methods for 1D datasets
- Scatter plots, correlation, and prediction for 2D datasets

Most people will have seen some, but not all, of this material. In my experience, it takes some time for people to really internalize just how useful it is to make pictures of datasets. I've tried to emphasize this point strongly by investigating a variety of datasets in worked examples. When I teach this material, I move through these chapters slowly and carefully.

Probability

This part covers:

- Discrete probability, developed fairly formally
- Conditional probability, with a particular emphasis on examples, because people find this topic counterintuitive
- Random variables and expectations
- Just a little continuous probability (probability density functions and how to interpret them)
- Markov's inequality, Chebyshev's inequality, and the weak law of large numbers
- A selection of facts about an assortment of useful probability distributions
- The normal approximation to a binomial distribution with large N

I've been quite careful developing discrete probability fairly formally. Most people find conditional probability counterintuitive (or, at least, behave as if they do—you can still start a fight with the Monty Hall problem), and so I've used a number of (sometimes startling) examples to emphasize how useful it is to tread carefully here. In my experience, worked examples help learning, but I found that too many worked examples in any one section could become distracting, so there's an entire section of extra worked examples. You can't omit anything here, except perhaps the extra worked examples.

The chapter on random variables largely contains routine material, but there I've covered Markov's inequality, Chebyshev's inequality, and the weak law of large numbers. In my experience, computer science undergraduates find simulation absolutely natural (why do sums when you can write a program?) and enjoy the weak law as a license to do what they would do anyway. You could omit the inequalities and just describe the weak law, though most students run into the inequalities in later theory courses; the experience is usually happier if they've seen them once before.

The chapter on useful probability distributions again largely contains routine material. When I teach this course, I skim through the chapter fairly fast and rely on students reading the chapter. However, there is a detailed discussion of a normal approximation to a binomial distribution with large N. In my experience, no one enjoys the derivation, but you should know the approximation is available, and roughly how it works. I lecture this topic in some detail, mainly by giving examples.

Inference

This part covers:

- Samples and populations
- Confidence intervals for sampled estimates of population means
- Statistical significance, including t-tests, F-tests, and χ^2-tests
- Very simple experimental design, including one-way and two-way experiments
- ANOVA for experiments
- Maximum likelihood inference
- Simple Bayesian inference
- A very brief discussion of filtering

The material on samples covers only sampling with replacement; if you need something more complicated, this will get you started. Confidence intervals are not much liked by students, I think because the true definition is quite delicate; but getting a grasp of the general idea is useful. You really shouldn't omit these topics.

You shouldn't omit statistical significance either, though you might feel the impulse. I have never dealt with anyone who found their first encounter with statistical significance pleasurable (such a person might exist, the population being very large). But the idea is so useful and so valuable that you just have to take your medicine. Statistical significance is often seen and sometimes taught as a powerful but fundamentally mysterious apotropaic ritual. I try very hard not to do this.

I have often omitted teaching simple experimental design and ANOVA, but in retrospect this was a mistake. The ideas are straightforward and useful. There's a bit of hypocrisy involved in teaching experimental design using other people's datasets. The (correct) alternative is to force students to plan and execute experiments; there just isn't enough time in a usual course to fit this in.

Finally, you shouldn't omit maximum likelihood inference or Bayesian inference. Many people don't need to know about filtering, though.

Tools

This part covers:

- Principal component analysis
- Simple multidimensional scaling with principal coordinate analysis;
- Basic ideas in classification;
- Nearest neighbors classification;
- Naive Bayes classification;
- Classifying with a linear SVM trained with stochastic gradient descent;
- Classifying with a random forest;
- The curse of dimension;
- Agglomerative and divisive clustering;
- K-means clustering;
- Vector quantization;
- A superficial mention of the multivariate normal distribution;
- Linear regression;
- A variety of tricks to analyze and improve regressions;
- Nearest neighbors regression;
- Simple Markov chains;
- Hidden Markov models.

Most students in my institution take this course at the same time they take a linear algebra course. When I teach the course, I try and time things so they hit PCA shortly after hitting eigenvalues and eigenvectors. You shouldn't omit PCA. I lecture principal coordinate analysis very superficially, just describing what it does and why it's useful.

I've been told, often quite forcefully, you can't teach classification to undergraduates. I think you have to, and in my experience, they like it a lot. Students really respond to being taught something that is extremely useful and really easy to do. Please, please, don't omit any of this stuff.

The clustering material is quite simple and easy to teach. In my experience, the topic is a little baffling without an application. I always set a programming exercise where one must build a classifier using features derived from vector quantization. This is a great way of identifying situations where people think they understand something, but don't really. Most students find the exercise challenging, because they must use several concepts together. But most students overcome the challenges and are pleased to see the pieces intermeshing well. The discussion of the multivariate normal distribution is not much more than a mention. I don't think you could omit anything in this chapter.

The regression material is also quite simple and is also easy to teach. The main obstacle here is that students feel something more complicated must necessarily work better (and they're not the only ones). I also don't think you could omit anything in this chapter.

In my experience, computer science students find simple Markov chains natural (though they might find the notation annoying) and will suggest simulating a chain before the instructor does. The examples of using Markov chains to produce natural language (particularly Garkov and wine reviews) are wonderful fun and you really should show them in lectures. You could omit the discussion of ranking the Web. About half of each class I've dealt with has found hidden Markov models easy and natural, and the other half has been wishing the end of the semester was closer. You could omit this topic if you sense likely resistance, and have those who might find it interesting read it.

Mathematical Bits and Pieces

This is a chapter of collected mathematical facts some readers might find useful, together with some slightly deeper information on decision tree construction. Not necessary to lecture this.

Urbana, IL, USA David Forsyth

http://www.springer.com/978-3-319-64409-7

Probability and Statistics for Computer Science

Forsyth, D.

2018, XXIV, 367 p. 124 illus., 84 illus. in color.,
Hardcover
ISBN: 978-3-319-64409-7